REVISION 1:

Observations on three-dimensional measurement of confined fission track lengths in apatite using digital imagery

Qingyang Li[1], Andrew Gleadow[1], Christian Seiler[1,2], Barry Kohn[1], Pieter Vermeesch[3], Andrew Carter[4], Anthony Hurford[3]

[1]School of Earth Sciences, University of Melbourne, Victoria 3010, Australia

[2]Now at Geoscience Australia, Box 378, Canberra, ACT 2601, Australia

[3]London Geochronology Centre, Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, United Kingdom

[4]Department of Earth and Planetary Sciences, Birkbeck College, Malet Street, London WC1E 7HX, United Kingdom

ABSTRACT

We report the results of a comparative study to explore the usefulness of 3D measurements of confined fission track lengths (TINTs) relative to horizontal confined track length measurements (dips ≤10°), and evaluate their suitability for thermal history modelling. Confined fission track lengths were measured in ten annealed Fish Canyon Tuff apatites containing synthetic mixtures of different length components, and two Durango apatites containing spontaneous fission tracks. Measurements were primarily carried out using a digital image-based microscope system, and compared to those from a regular optical drawing tube-digitizing tablet set-up and a confocal laser scanning microscope. The results indicate that 3D measurements of confined track lengths are closely comparable to

conventional horizontal track measurements, and the mean track lengths of inclined (dips >10°) and horizontal (dips ≤10°) confined tracks from the one sample are equivalent within the measurement uncertainty. A strong dip-bias was observed, so that almost all the confined tracks measured were dipping at <30°, and the great majority (~70%) were dipping at ⩽10°, thereby qualifying as 'horizontal' confined tracks. Our results suggest that a useful increase of more than 40% in sample size can be achieved from including dip- and refraction-corrected 3D track length measurements. Some evidence was seen for a small bias in favor of shorter tracks at higher dip angles but this has very little influence on the mean lengths or length distributions up to the practical limit of dips (~30°) observed in these measurements. Results obtained using the same measurement system by a single analyst over time, and between six different observers in the one laboratory, show good reproducibility. These results also agree well with conventional horizontal confined track length measurements in the same samples in the second laboratory involved. We conclude that 3D measurements of confined track lengths, including both horizontal and inclined tracks, are suitable for use in current fission track annealing models derived from experiments using horizontal confined tracks.

Keywords: Thermochronology, fission track dating, apatite, confined track lengths, 3D measurement, digital imaging

INTRODUCTION

Apatite fission track (AFT) thermochronology is used for reconstructing geological thermal histories through combining apparent age and confined track length measurements (e.g.

43 Gleadow et al. 2002; Gallagher 2012; Ketcham 2005). Fission tracks form continually over
44 time, but the length of each track is subjected only to the subsequent thermal history since its
45 formation. Thus, the distribution of confined track lengths in a particular sample is
46 characteristic of its thermal history (Gleadow et al. 1986) since entering the partial annealing
47 zone. Detailed thermal histories can be reconstructed from the combined fission track length
48 and age data by using fission track annealing models (e.g. Ketcham et al., 1999; Laslett and
49 Galbraith 1996; Laslett et al. 1987).

50 Laslett et al. (1982) pointed out that all practical schemes for sampling etched fission track
51 lengths will be subject to various kinds of bias. They concluded that sampling horizontal
52 confined fission tracks will be the least biased and provide the closest approximation to the
53 underlying distribution of unetched track lengths. Since that work, and later empirical studies
54 by Gleadow et al. (1986), standard practice has been to measure the projected lengths of such
55 horizontal confined track (HCTs). In reality, ‘horizontal’ is taken to mean tracks dipping at up
56 to ~10° (Donelick et al. 2005) or even ~15° (Laslett et al. 1982), for which the resulting
57 errors introduced by measuring only the horizontal length component are relatively small,
58 ~1.5 % to ~3.4% respectively.

59 In the absence of actual dip measurements, it is obviously difficult to apply these criteria for a
60 particular track to be horizontal in any rigorous sense. Mostly the operator makes a
61 qualitative judgment based on the focus and appearance of the track in transmitted and/or
62 reflected light, which will therefore depend to some extent on the microscope conditions and
63 the level of experience. This judgment is made more complex by the fact that tracks in apatite

64 appear to dip at significantly less than their true dips due to refraction effects when observed
65 under air (Laslett et al. 1982).

66 Another consideration from those early studies was that, while it was simple to measure the
67 horizontal component of a track length, it was more difficult and laborious to measure the
68 vertical component required for measuring the dip with comparable precision, using typical
69 microscopes available at that time. Such measurements were clearly possible (e.g. Dakowski,
70 1978), but have become much more straightforward and convenient with the current
71 generation of fully motorized and digitally controlled microscopes (e.g. Gleadow et al, 2015).
72 A significant limitation caused by using only HCTs, however, is that this restriction reduces
73 the potential sample size for measurements on features that are already rare events. In the
74 case of young or low-uranium apatites, it is often difficult to locate enough HCTs to make
75 thermal history modeling possible.

76 Measurement errors introduced by including tracks that are not strictly horizontal are likely to
77 be small in most cases, but in standard practice will always be present, and potentially to
78 different degrees between samples and observers. Such errors could contribute to some
79 degree to the poor inter-laboratory reproducibility of HCT length data reported by Ketcham
80 et al. (2009; 2015). In principle, it should always be better to correct for the dip and measure
81 the true etched lengths of confined tracks in three dimensions, even for HCTs. Jonckheere
82 and Ratschbacher (2010) reported an approach to measuring the lengths of non-horizontal
83 tracks in apatite with the potential to significantly increase the available sample size, and
84 pointed out that the more complex biases inherent in such measurements should not introduce

insurmountable problems. It is not the purpose here to evaluate the range of potential biases present, however, but rather to explore empirically just how significant their cumulative effect might be relative to standard HCT measurements. The primary aim is therefore to test the usefulness and practicality of including 3D measurements of non-horizontal track lengths in routine fission track analysis.

The measurements reported here utilize a fully motorized microscope system to capture multi-plane image sets (so-called z-stacks) at precisely controlled vertical intervals to digitally image the 3D structure of the etched fission tracks. A secondary aim of this study is therefore to establish the degree to which such image-based techniques produce measurements that are comparable to those from older optical drawing-tube systems attached to the microscope. The confined fission track lengths acquired using an image-based system, (here referred to simply as '3D lengths') are automatically corrected for both dip and refraction, and represent the 'true' lengths of the etched tracks. The orientation of each track relative to the crystallographic c-axis is also automatically determined in this system. Including additional lengths for what are here termed inclined confined tracks, i.e. those with dip angles of >10°, can increase the sample size.

3D track length measurements are reported from twelve apatite samples, containing fission tracks with length distributions of varying degrees of complexity. We compare 3D track length measurements of dipping tracks with conventionally measured HCTs, and investigate the difference between horizontal and inclined track length measurements. We also examine whether the precision of 3D measurements can be improved by using more closely spaced

image stacks, and by using confocal laser scanning microscopy with its inherently higher image resolution (e.g. Petford and Miller 1992, 1993). Finally, we evaluate the consistency of 3D length measurements by comparing results obtained by a single analyst over time on several samples, and by different observers on one particular sample that was also part of the comparative study by Ketcham et al. (2015).

LENGTHS AND ORIENTATIONS OF CONFINED FISSION TRACKS

The length of a fission track at any arbitrary orientation can be calculated from a simple set of geometric equations using a coordinate system that is defined relative to an observation surface and the crystallographic orientation (Fig. 1, after Galbraith and Laslett, 1988). The observation surface in an apatite grain is normally selected to be parallel to the crystallographic *c*-axis. The coordinate system consists of a plane XY that is parallel to this observation surface, where the X-axis is parallel to the crystallographic *c*-axis. The 3D length of a confined track l_t orientates from the origin O with dip angle θ, and has a projected length l_p on the observation plane XY and a depth D parallel to the Z-axis. The angle between l_t and the *c*-axis is denoted by φ (Galbraith and Laslett 1988), while ω represents the azimuth angle in the observation plane between l_p and the *c*-axis.

Once the XYZ coordinates of the end points of a confined track are known, l_t, θ, and φ can be calculated from l_p, D, and ω using the following equations (after Jonckheere and Ratschbacher, 2010), assuming that the polished surface is perfectly flat:

$$l_t = \sqrt{{l_p}^2 + D^2} \qquad (1)$$

$$\theta = cos^{-1}\frac{l_p}{l_t} \qquad (2)$$

$$\varphi = cos^{-1} \frac{l_p \cos \omega}{l_t} \quad (3)$$

Because of the contrasting refractive indices between apatite and the surrounding air, the apparent depth of the end point of the track (d_a in Fig. 1) is shallower than the true depth (d) when observed using a dry objective lens. The apparent 'true' length of the tracks is given by l_a in Fig. 1. The true depth is determined simply as the product of the apparent depth and the refractive index of apatite. While the difference between the ordinary and extraordinary rays in apatite is at its maximum on the prismatic observation surfaces used, the birefringence is so low that this can be ignored and an average refractive index applied. Here a refractive index of 1.634 was used, as a reasonable average for near fluorapatites (e.g. Deer et al., 1965, p 507).

This refraction effect can usually be ignored in projected track length measurements (typical HCTs), as the effect is small in this case, but can become significant when tracks that are not strictly horizontal are included. If uncorrected, this refraction effect means that a track which appears to dip at 10° is actually dipping at 16° and the 1.5% measurement error increases to 4%, whereas an apparent dip of 15° is actually 25°, for which the measurement error would be nearly 10%. In this study, the criterion for tracks to be 'horizontal' is taken to be ≤10° true dip after correction for the refractive index.

Most of the comparisons made here will be between different mean track lengths, with their respective standard errors (SE) and standard deviations (SD) of the distributions. When comparing confined tracks over different dip ranges, the terms HCT and ICT will be used to denote mean lengths of 'horizontal' and 'inclined' confined tracks dipping at ≤10° and > 10°

148 respectively. We will also refer to mean lengths of ‘all’ confined tracks (ACT) over the full
149 range of dips observed, and will differentiate between ‘projected’ lengths (i.e. projected onto
150 the XY plane in Fig. 1) and ‘true’ lengths, corrected for dip, by using the subscripts ‘p’ and ‘t’
151 respectively.

152 ## SAMPLES AND METHODS

153 ### Sample details

154 A total of 12 samples from two well-known apatite reference materials, the Fish Canyon Tuff
155 (FCT, 10 samples) and Durango (DUR, 2 samples) apatite, were used in this study. The Fish
156 Canyon Tuff samples were prepared at the London Geochronology Centre at University
157 College London (UCL) and consisted initially of twelve aliquots of separated apatite. All
158 were first annealed at 600ºC for 24 hours to remove all pre-existing spontaneous fission
159 tracks. The aliquots were then irradiated in the former HIFAR Reactor at Lucas Heights,
160 Australia, with a total neutron fluence of 9×10^{15} n/cm^2 to induce ^{235}U fission tracks. One
161 aliquot, containing only fresh induced tracks with a mean track length of ~16 µm, was set
162 aside at this point as Control 1. Splits of the remaining 10 samples were reheated for 1 hour
163 to temperatures of 300°, 350° or 370°C to produce different degrees of partial annealing to
164 mean track lengths of ~13, ~11 and ~8 µm represented by Controls 2, 3 and 4 respectively.
165 The sample Control 2 (~13 µm) was not available to this study and is not considered further.
166 The remaining seven FCT samples were then re-irradiated in the same reactor with various
167 thermal neutron fluences between 1.1×10^{15} and 2.5×10^{16} n/cm^2 to produce two-component
168 mixtures of one of the three annealed components and a new unannealed (~16 µm)

169 component in varying proportions as indicated in Table 1. This sample suite has other
170 important calibration applications, but is used here purely to present a known range of track
171 length distributions and mixtures of known length components.

172 Sample DUR was prepared at the Melbourne Thermochronology laboratory and consists of
173 suitably sized crystal fragments (80~200 μm) that were obtained by crushing a single
174 Durango apatite crystal. The sample was analyzed in its natural state, containing only
175 spontaneous ^{238}U fission tracks. Sample DUR-4 is a sliced ~1 mm thick plate of Durango
176 apatite that was part of a previous inter-laboratory comparative fission track length study
177 (Ketcham et al., 2015) and was prepared at UCL. The crystal was cut parallel to the *c*-axis
178 before being heated to 500°C for 24 hours in order to fully anneal all spontaneous tracks. It
179 was then irradiated at the same reactor to generate induced tracks, before being reheated
180 again to 240°C for 10 hours in order to reduce tracks to lengths that would be expected in a
181 rapidly cooled, natural volcanic sample.

182 All apatites were mounted in epoxy resin on glass slides, ground and polished to expose
183 internal surfaces. At this stage, ^{252}Cf fission tracks were implanted in the surface of samples
184 DUR and DUR-4 in order to increase the number of confined track lengths for measurement
185 (Donelick and Miller 1991). Etching conditions for all mounts were identical: 5M HNO_3 for
186 20 seconds at 20°C. FCT samples were mounted and etched at UCL, while DUR and DUR-4
187 were prepared at the University of Melbourne (UoM).

188 **Length measurements using conventional wide field microscopy**

189 Confined track lengths where measured by Analyst A from UCL and Analysts 1-6 from the

UoM group using different equipment and techniques. In both cases, mineral mounts were set-up under the microscope and referenced to the stage coordinate system. The operator then scanned grain mounts for confined tracks etched though surface-intersecting tracks (Track-in-Track features - TINTs; Lal et al. 1969, Green 1981, Donelick et al. 2005), marking the location of suitable tracks for analysis. The criteria for track selection were to only include those TINTs with distinct track ends, excluding tracks with blurred ends due to overlap with other features, or with thin or faint track ends that may not have been fully etched (Laslett et al. 1984). All measurements were made on prismatic surfaces of apatite grains with their *c*-axes in the plane of the microscope stage, as determined by observation under circular polarized light, or the sharpness of polishing scratches and parallel orientation of the long-axes of surface etch pits (D_{par}; Donelick et al. 1999; Green et al. 1986; Gleadow et al. 2009a).

At UCL, horizontal confined track lengths were measured with a Zeiss *Axioplan* microscope using a 100x dry objective (total magnification 1250x), and a Calcomp Drawing Board III tablet with an LED attached to the cursor. The LED light is projected via a drawing tube into the microscope field of view where it produces a bright spot ~0.2 μm in diameter. The projected length of each track was determined by clicking the cursor at each end. Only tracks with estimated dips <15° (apparent) were measured. Calibration of the digitizing tablet used a certified stage micrometer with 2 μm divisions. Precision of the measuring system is estimated at ± 0.11μm.

At the UoM laboratory, track length measurements were carried out as refraction and

211 dip-corrected 3D lengths using an automated image acquisition and processing system
212 developed in house, which consists of a motorized Zeiss *Axio-Imager* M1m microscope
213 controlled by the *TrackWorks* software package (Gleadow et al. 2009a, 2009b; 2015;
214 Gleadow and Seiler 2015). The microscope is fitted with a motorized stage system with
215 vertical movements in 25 nm steps. For each track located by the operator, *z*-stacks of digital
216 images, at vertical spacings of 0.1, 0.2 or 0.3 μm between image planes, were captured
217 autonomously in transmitted and reflected light at previously marked locations. All images
218 were captured using a 100x dry objective and a 3.3 Megapixel Zeiss ICc3 camera on a 0.5x
219 C-mount adapter. At the time of acquisition the image sets were automatically cropped to a
220 box 35×35 μm in area around the location of each identified confined track. The captured
221 image sets were then archived to a local network storage array.

222 Archived fission track image sets were later retrieved and analyzed on a computer using the
223 fission track image analysis and measurement software *FastTracks* (Gleadow et al. 2009b;
224 2015; Gleadow and Seiler 2015). Track lengths were measured by focusing through the
225 digital image stacks on the monitor and clicking at each end of the confined track with the
226 cursor at a total effective magnification of ~6,000x - 10,000x. The point of this higher
227 magnification is not that it contains any further information than may be observed under the
228 microscope, but that it minimizes placement errors in positioning the cursor at the ends of a
229 track. Image scale was calculated to be 0.069 μm/pixel, based on total magnification to the
230 camera and the pixel spacing in the Sony CCD sensor (3.45 μm), and confirmed by direct
231 calibration against a Pyser-SGI S21 stage micrometer with divisions of 10 μm.

At least 170 confined tracks were located, imaged and measured on each sample, except for FCT G, where only 120 tracks could be found in the entire sample (Table 2). For FCT samples, an initial sampling of ~100 tracks were selected for measurement irrespective of their dip angle θ. A second selection was then added until at least 100 HCTs were included in the measurement to make that component comparable to most conventional measurements. For samples DUR and DUR-4, track lengths were measured from 355 and 223 captured track image sets, respectively, each containing one or more confined tracks. The particular selection strategy for DUR-4 is described in more detail below under 'Reproducibility between multiple analysts'.

All images were captured on *c*-axis parallel surfaces, and the *c*-axis direction in this surface plane was determined automatically by image analysis from the mean direction of the long-axes of the parallel track openings in reflected light, i.e. the orientation of the D_{par} parameter. These automatically determined azimuth directions were adjusted manually if required. In addition to the calculated true and projected confined track lengths, the orientation angles θ and φ were also automatically determined using *FastTracks*. Analysts at UCL and UoM were unaware of each other's results during analysis.

Confocal laser scanning microscopy (CLSM)

Confocal laser scanning microscope measurements were made using a Zeiss *LSM700* materials science module utilizing a single 405 nm laser attached to a Zeiss *Axio Imager* Z1m microscope, controlled by Zeiss *ZEN* software. The fully motorized Z1m microscope is fitted with a piezo drive x-y scanning stage and motorized z-axis with a vertical resolution of 10

nm, monitored by a piezo encoding device.

CLSM length measurements were carried out on exactly the same tracks that had previously been selected and analyzed by conventional wide-field microscopy. Analysis was only carried out on the FCT samples as these cover a wide range of possible track length distributions. During capture, step sizes of the z-stack and objective magnification were the same as that used during the conventional microscopy.

RESULTS AND DISCUSSION

Horizontal and 3D confined fission track lengths

A comparison of the mean track length results from Analyst 1 at UoM and Analyst A at UCL is shown in Table 2 and individual track length (l_t) distributions for Analyst 1 are shown in Fig. 2. The 3D measurements of Analyst 1 for mean tracks lengths of confined tracks over all dip angles (ACT_t) range from 15.89(05) to 8.25(18) (±1 SE), from the unannealed Control 1 to the most highly annealed Control 4 sample. The spontaneous tracks in DUR and DUR-4 both contain a single length component with mean track lengths 14.13(05) and 14.24(06) µm, consistent with previous measurements on Durango apatite (e.g. Gleadow et al. 1986; Green 1988; Kohn et al. 2002). The FCT samples A-H have mixed length distributions with means from ~12-15 µm, which shorten in line with the ratio of unannealed to annealed track lengths, and the different components can be seen in the complex length distributions in Fig. 2. The distribution for DUR-4 is essentially identical to DUR and not illustrated.

The mean projected lengths of horizontal confined tracks (HCT_p) measured at UCL and UoM

agree closely, but the UCL lengths tend to be slightly longer in most samples, as shown in Fig. 3A. The differences ranged from 0 to 1.22 μm, with an average 0.24 μm. Only the means for Control 1 were significantly different at the 95% confidence level. These discrepancies probably include typical inter-laboratory factors such as differences in microscope configuration, system calibration and observer biases, but they are very small compared to the range observed between different laboratories reported by Ketcham et al. (2015) and Barbarand et al. (2003). The two largest deviations of 0.39 and 1.22 μm were found for samples FCT F and G, both of which show complex two-component length distributions with significant numbers of short tracks. This suggests that in these two cases differences in track selection are the most significant factor.

Very little difference was observed between measurements of the projected lengths of horizontal tracks (HCT_p) and the true, dip- and refraction-corrected, 3D lengths (ACTt) of confined tracks in all samples measured by Analyst 1 at UoM (Table 2, and Fig. 3B). These measurements were all made by the same observer using an identical measurement system and the mean lengths for tracks at all dips tend to be slightly shorter than the means projected lengths of horizontal tracks. The differences in this case range from -0.18 to 0.20 μm with a mean of -0.02 μm, and none are significant at the 95% confidence level. Once again, the two largest deviations are for FCT F and G, again suggesting that small differences track selection between the two length peaks have been a factor, although in this case, the effect is very small.

The true 3D length measurements were further divided into horizontal ($\theta \leq 10°$) and inclined

294 (θ >10°) components to compare the mean lengths of the shallow (HCT_t) and more steeply
295 dipping (ICT_t) fractions. A summary of mean track length results for the two sub-groups is
296 shown in Table 3 and the differences between them in Fig. 3C. The results show that the
297 mean track lengths are systematically longer for horizontal than for inclined tracks in nearly
298 all cases, except for two most complex samples (FCT G and FCT H) where the difference is
299 reversed. The differences range from -0.13 to 0.82 μm with a mean value of 0.28 μm. Except
300 for the two extreme samples, where track selection is again likely to be the dominant factor,
301 the differences are small but consistently in one direction. This implies that 3D measurements
302 for confined tracks at higher dip angles are systematically shorter than those at lower dips,
303 but in most cases the differences between the means are still not significant at the 95%
304 confidence level.

305 **Orientation analysis of individual 3D lengths**

306 Individual 3D confined track length measurements l_t are plotted against dip angles θ in Fig. 4.
307 The alignment of the measurements into sub-parallel 'dot-curve' arrays is due to the depth
308 measurements being quantized by the discrete layer planes sampled in the image z-stacks, as
309 illustrated in Fig. 5. Each of these dot-curves represents tracks observed to terminate in the
310 same image depth plane.

311 The four samples containing a single unannealed length component (Control 1, DUR), or
312 once-annealed fission tracks (Controls 3 and 4) show a relatively uniform distribution of
313 individual track lengths with increasing dip angle. The dispersion of lengths in each
314 component increases with the degree of annealing, as expected, due to the anisotropic

315 annealing of tracks in different crystallographic orientations (e.g. Green et al. 1986). All of
316 the two-component mixtures (FCT A-H) show fields of lengths belonging to each component,
317 with varying degrees of overlap depending on the degree of annealing. For samples
318 containing the most strongly annealed components, with means of ~11 and ~8 μm, the two
319 length components are essentially separate from each other (Fig. 4: FCT B and C, and
320 especially FCT F, G, H).

321 In most of the observed length components the maximum individual track lengths tend to
322 decrease, and the minimum lengths increase, with increasing dip angle, giving a tapering
323 field towards the higher dip angles. This is most obvious for the longer length components,
324 where the density of the field is greatest. Perhaps surprisingly, almost all of the measured
325 tracks lie at dips of 30° or less, with only a few outliers beyond this and only three beyond
326 40°. This may be due to the difficulty of identifying confined tracks at higher dips given the
327 limited depth of focus of the microscope at the high magnification used. There is a tendency
328 for the few tracks dipping at >30° to be shorter than the main group of HCTs at <10°, and it is
329 likely that these extreme outliers are lowering the mean lengths for ICTs to the small degree
330 observed in Table 3 and Fig. 3C.

331 The histogram of all dip angles in Fig. 4B shows even more strongly how the number of
332 tracks sampled decreases rapidly with increasing dip, so the apparent narrowing of the field
333 of lengths towards higher dip angles might reflect the more limited sampling in this region.
334 About 70% of the observed tracks dip at ≤10°, and would therefore qualify as HCTs in a
335 conventional measurement. Even more (~90%) would be HCTs if the threshold were set at

15°. As indicated previously few confined tracks were observed at dip angles above ~30º and almost none beyond 40°.

These results suggest that there is a strong real or observational bias towards low angle tracks in 3D measurements of confined fission tracks. This dip-bias is probably due to the limited depth range over which the confined tracks are sampled. All the measurements in this study were made on track-in-track (TINT) features, where the lengths of the surface intersecting semi-tracks, from which they are etched, limits the intersection depth from which they can be revealed. This depth will range up to the maximum length of a confined track and will on average be about half of this length. Confined tracks that are intersected only a few micrometers below the surface must of necessity be almost horizontal because if the dip was greater they would intersect the surface and no longer qualify as confined tracks. Longer tracks at higher dip angles, must therefore be intersected further below the surface, on average, and might therefore be etched to a slightly lesser degree because of the finite time taken for the etchant to reach their ends. This might be another factor in the very slightly reduced length apparent for ICTs, compared to HCTs observed in Table 3.

In principle, it might be expected that shorter tracks would be observed to higher dip angles than longer tracks, as the longer tracks would be more likely to intersect the surface and so be excluded. This has been termed a surface-proximity bias by Galbraith (2005 p.157). However, the results for the mixed length components in Fig. 4(A) do not show any consistent trend in this regard, with both long and short track groups occurring over a similar angular range. It would appear then that this postulated proximity-bias is not a significant factor limiting the

use of 3D length measurements, at least over the limited dip range that is actually sampled in practice.

Sensitivity to step-size in the image stack

The quality of an image-based measurement of a fully etched fission track depends on the resolution of the input image in all three dimensions, as well as the precision with which the track ends can be defined. The image resolution is controlled by the magnification, the numerical aperture of the optics and the wavelength of the light source, as well as the pixel resolution of the image sensor. In our experiments, images were captured digitally at a pixel resolution of 70 nm, which exceeds the diffraction-limited resolution of the microscope optics for visible light (~280 nm with a 100x dry objective, numerical aperture NA=0.9). This satisfies the Nyquist limit in the image plane (~100 nm), which defines the sampling rate required to faithfully digitize an analog signal. The same does not apply to the z-direction, however, where the image spacing of typically ~300 nm (or ~490 nm after correction for the refractive index) exceeds the optical resolution in this direction and the stack is under-sampled. The effect of this limitation was tested first by reducing the step-size between image planes within the captured z-stacks until they were close to the Nyquist limit (~100 nm), and second by utilizing confocal laser-scanning microscopy, which achieves a higher resolution than is possible with conventional wide field microscopy.

The step-size of an image stack refers to the vertical interval between captured image planes of the stack, which is used to calculate the true (refraction-corrected) vertical distance between the two ends of a fission track when they are in focus. Reducing the step-size

increases the vertical sampling of images, thereby increasing the vertical resolution of the stack so that track ends can be measured more precisely, but at the expense of larger stack sizes.

To evaluate the effects of the step-size, repeat 3D length measurements were made on images of the same set of confined fission tracks in sample DUR captured using three different step-sizes: 0.1, 0.2 and 0.3 μm (corresponding to refraction-corrected step-sizes of 0.16, 0.33 and 0.49 μm). Measurements were made on totals of 105, 106 and 117 confined tracks respectively, but most of the tracks were common to two or three of the image sets. Results are shown for mean lengths of the 80 tracks common to all three step-sizes in Table 4, for all measurements in Fig. 6A and for the 88 individual lengths common to the 0.1 and 0.3 μm step-sizes in Fig. 6B. The maximum difference between the mean lengths is 0.06 μm, and none are statistically significant. The mean track lengths (ACT_t) for the three step-sizes are essentially identical and show no systematic difference between the minimum and maximum increments.

Individual 3D track lengths plotted against dip angles for these three step-sizes are shown in Fig. 6A. The spacing between the dot-curves decreases with decreasing the step-size, but there is no systematic change in the overall field covered. A comparison of all paired lengths on the same tracks obtained by the largest (0.3 μm) against smallest (0.1 μm) step-sizes (Fig. 6B) shows that data points are tightly scattered around the 1:1 line (root mean square deviation: 1.6%). This finding is consistent with the small difference in the mean track lengths (<0.03 μm) arising from the three different step-sizes (Table 4), suggesting that the differences are negligible.

Confined track length measurements by Confocal Laser Scanning Microscopy

Track length measurements of FCT samples were repeated using CLSM in order to assess whether the increased resolution of CLSM is advantageous for 3D confined track length measurements (c.f. Petford and Miller, 1992, 1993). As can be seen in Fig. 7, the same tracks imaged using CLSM are more sharply resolved and the track ends better defined than in a conventional wide field image. Track lengths on the same individual tracks measured by both CLSM and conventional optical microscopy are highly correlated and essentially identical (Fig. 8). The mean track lengths, standard deviation and standard errors for all the samples obtained by both methods are in close agreement with each other and essentially indistinguishable within error (Table 5). On average, the CLSM lengths are very slightly longer than 3D lengths by wide field microscopy (by 0.10 μm), but the difference is insignificant (root mean square deviation: 0.94%).

A key restriction of CLSM, however, is that measurements can only be carried out on tracks with relatively low dip angles (Petford and Miller, 1993). The mean θ of tracks measured on the CLSM is between 2.4° to 4.1°, with the steepest track dip measured at 15.7° (Table 5). That is because confocal laser imaging is inherently an incident light method and only tracks with low θ reflect sufficient light to enable them to be detected and measured. In an effort to overcome this limitation, we attempted to enhance the visibility of inclined tracks using a fluorescent dye and captured images in fluorescence mode, but the number of measurable tracks did not increase due to the resulting overall poor illumination conditions.

Reproducibility of 3D measurements over time

421 Repeat measurements were made by Analyst 1 on the same captured image sets after an
422 interval of ~2.5 years (Table 6) to assess the reproducibility of these results. Analyses were
423 made on five of the FCT and DUR samples and the measurements were made on exactly the
424 same tracks in most cases, although slightly more or less confined tracks were judged suitable
425 in three of the samples (Table 6). Thus the selection of tracks was essentially identical for
426 both cases and the only differences were in the measurements themselves. In all cases the
427 replicates closely reproduce, and are statistically indistinguishable from, the original
428 measurements. The initial measurements are very slightly, but consistently, higher than the
429 second, with differences ranging from 0.03-0.12 μm, and a mean of 0.07 μm. On the other
430 hand, the standard deviations from the repeat analyses are consistently slightly greater, with
431 differences ranging from 0-0.12, with mean of 0.04 μm, but none of these differences are
432 statistically significant. The reason for these slight systematic differences is attributed to a
433 change in the magnification used on the monitor, leading to a subtle difference in defining the
434 ends of tracks. However, the differences are insignificant and all the track length
435 measurements are highly reproducible.

436 **Reproducibility between multiple analysts**

437 In order to assess the reproducibility of 3D length measurement results between different
438 analysts, six experienced analysts in the UoM laboratory were requested to measure confined
439 tracks on a set of archived images captured from sample DUR-4. This sample was part of the
440 blind inter-laboratory comparison experiment reported by Ketcham et al. (2015). The
441 measurements reported here were carried out before that study was published, so the

442 comparative results were unknown at the time of measurement. To allow for an element of
443 individual selection of the tracks to be measured, a total of 400 TINTs were identified across
444 223 locations in the mount. Image stacks were acquired from all locations and distributed to
445 the analysts. Each analyst was asked to select and measure at least 100 tracks from the entire
446 set, based on personal criteria as to which tracks were satisfactory for measurement. The
447 results are presented in Table 7 and Fig. 9. The latter also includes the comparative data from
448 Ketcham et al. (2015), acquired from identically prepared apatite samples and measured by
449 55 analysts in 30 different laboratories.

450 Mean 3D lengths in DUR-4 from this study ranged from 14.12 to 14.29 μm between the
451 different analysts with a mean of 14.20(03) μm (SE), and SDs ranged from 0.79 to 0.98 μm.
452 The consistency of these measurements is excellent and even the maximum difference
453 observed (0.17 μm) is not significant at the 95% confidence level. The mean value is also
454 consistent with the mean of the international comparison. The variability is substantially
455 lower than most of the measurements on the same sample from other laboratories in Ketcham
456 et al. (2015) at both the inter-laboratory and intra-laboratory level (Fig. 9).

457 **IMPLICATIONS**

458 'Horizontal' confined track length measurements, which are a central component of apatite
459 fission track thermochronology, in reality include tracks with a range of dips up to a
460 threshold of usually ~10°. The resulting length measurements are projected lengths that will
461 mostly be shorter than the true lengths by a small, and presumed negligible, amount. The
462 discrepancies will increase with increasing dip angle, however, and this is compounded by

refraction in apatite, which makes track dips appear significantly less than they actually are. Where dips are not measured explicitly, it is difficult to rule out the possibility that tracks significantly above the dip threshold are being included in the measurement. Differing dip thresholds for apparently ‘horizontal’ tracks could therefore explain at least some of the previously observed variability between analysts, although the overall effect is likely to be relatively minor in most cases.

3D measurements of confined fission track lengths based on captured z-stack images, were used in this study to determine the true track lengths, corrected for both dip and refractive index, thereby overcoming the small but known errors associated with projected length measurements. Our observational results across apatite samples with a wide range of track length distributions, show that 3D length measurements are actually closely comparable to the commonly used ‘horizontal only’ projected track length measurements (Table 2, Fig. 3B). These 3D measurements showed excellent reproducibility between individual analysts, and between replicate measurements over time. In addition, the image-sets upon which they are based provide a permanent digital record of those measurements, which could assist in standardization between laboratories.

A major contributing factor to the very close agreement between the dip-corrected and projected measurements of confined track lengths is the presence of a strong dip-bias in the 3D measurements favoring shallow dipping tracks. The great majority of confined tracks in the 3D measurements (~70%) are thus dipping at low angles (≤10° true dip) and meet the criterion to be measured as HCTs. This natural control on the observed dip range is most

484 probably due to the limited depth over which TINTs can be sampled from
485 surface-intersecting semi-tracks.

486 Our results also reveal a variable tendency for the mean length of inclined tracks dipping
487 at >10° to be slightly shorter than those for horizontal tracks (dips ≤10°) in most samples, by
488 an average of ~0.3 μm. This probably reflects a small bias in favor of shorter confined tracks
489 at higher dip angles, where longer tracks might intersect the surface and therefore be
490 excluded. Such a surface proximity-bias is not obvious, however, in the proportions of short
491 and long tracks at higher dip angles in the individual track data in Fig. 4, but this does not
492 rule out a small effect on the mean 3D length. Other factors, such as a possible lesser degree
493 of etching for tracks with steeper dips due to their deeper location in the crystal, might also
494 be involved in the small differences observed. However, these are likely to be subordinate to
495 the influence of other factors, such as the selection of tracks for measurement. It is possible
496 that any small deficit in the 3D lengths at high dip angles from these causes may actually
497 contribute to the concordance of dip-corrected 3D measurements and projected HCT lengths,
498 which are also subject to a very small underestimation of the true lengths due to the
499 uncorrected dips.

500 One of the potential benefits of using 3D length measurements is the expected increase in the
501 number of tracks available for measurement, and the results reported here show this increase
502 to be typically about 40%. This increase was less than anticipated, and substantially less than
503 the 3-4 times increase reported by Jonckheere and Ratschbacher (2010) using much
504 deeper-penetrating implanted heavy ion tracks. The reason for this difference is probably that

505 the measurements reported here were made on confined TINTs etched from mostly relatively
506 short semi-tracks allowing etchant penetration from the surface, which means that almost all
507 measured tracks were at dips of <30°, and the great majority were dipping at <10°.

508 Sensitivity studies indicate that there is little potential to improve the quality and consistency
509 of the 3D length results either by reducing the image spacing in the captured image z-stacks,
510 or by using Confocal Laser Scanning Microscopy to increase the optical resolution. The
511 statistically identical results obtained on the same tracks by conventional optical microscopy
512 and CLSM mean that the latter has no advantages for track length measurement, and is
513 impractical for routine use.

514 A distinct advantage of a digital image-based 3D measurement method is that it enables more
515 consistent application of measurement protocols, which is realized principally by allowing
516 for more precise cursor placement at greatly enlarged magnification. Digital image sets can
517 also be shared easily between laboratories, providing an additional aid to standardization of
518 procedures. However, it is also clear that strict control of other factors, such as etching,
519 equipment conditions, sampling criteria, system calibration, etc., remain important in the
520 effort to enhance compatibility of track length data sets. We suggest that including
521 image-based 3D measurements with other endeavors to standardize measurement procedures,
522 can contribute to the ongoing efforts to improve the reproducibility of track length data across
523 different laboratories.

524 The results of this study indicate that 3D confined track length measurements on TINTs are
525 directly comparable to conventional ‘horizontal only’ track length measurements and lead to

a moderate increase in the number of tracks available. As a result, it is concluded that 3D confined track length measurements should be compatible with current annealing models based on horizontal track length measurements, and therefore useful for thermal history reconstruction. The moderate increase in the number of tracks available for measurement in this way may be particularly useful for samples that lack sufficient horizontal tracks for robust thermal history modeling.

ACKNOWLEDGEMENTS

This work has been supported by infrastructure grants through the *AuScope* Program under the Australian National Collaborative Research Infrastructure Strategy (NCRIS). Development of the Automated Fission Track Analysis System at the University of Melbourne was initially supported through an Australian Research Council Linkage Project grant (LP0348767) in partnership with *Autoscan Systems Pty Ltd*. We would like to thank Ling Chung and Abaz Alimanovic for technical support in the laboratory, and Ling Chung, Vhairi Mackintosh, Matthew Barrand, Guangwei Li, Zhiyong Zhang, and Jianhui Liu for assistance with the project. The manuscript was greatly improved by detailed reviews from Raymond Jonckheere and Rich Ketcham, which were very much appreciated.

REFERENCES

Barbarand, J., Hurford, T., and Carter, A. (2003) Variation in apatite fission track length measurement: implications for thermal history modeling. Chemical Geology, 198, 77-106.

547 Dakowski, M. (1978) Length distributions of fission tracks in thick crystals. Nuclear Track
548 Detection, 2, 181-189.
549 Deer, W.A.,Howie, R.A., and Zussman J. (1966) An introduction to the rock forming
550 minerals. Longman, Green and Co., London, 528p.
551 Donelick, R.A., and Miller, D.S. (1991) Enhanced TINT fission track densities in low
552 spontaneous track density apatites using ^{252}Cf-derived fission fragment tracks: A model
553 and experimental observations. International Journal of Radiation Applications and
554 Instrumentation. D. Nuclear Tracks and Radiation Measurements, 18, 301-307.
555 Donelick, R.A., Ketcham, R.A., and Carlson, W.D. (1999) Variability of apatite fission track
556 annealing kinetics: II. Crystallographic orientation effects. American Mineralogist, 84,
557 1224.
558 Donelick, R.A., O'Sullivan, P.B., and Ketcham, R.A. (2005) Apatite Fission track Analysis.
559 Reviews in Mineralogy and Geochemistry, 58, 49-94.
560 Gallagher, K. (2012) Transdimensional inverse thermal history modeling for quantitative
561 thermochronology. Journal of Geophysical Research 117, B02408
562 Galbraith, R. F. (2005) Statistics for fission track analysis. Chapman & Hall/CRC, Boca
563 Raton, 219 p.
564 Galbraith, R.F., and Laslett, G.M. (1988) Some Calculations Relevant to Thermal Annealing
565 of Fission Tracks in Apatite. Proceedings - Royal Society. Mathematical, Physical and
566 Engineering Sciences, 419, 305-321.

Gleadow, A.J., and Seiler, C. (2015) Fission Track Dating and Thermochronology. *In* Rink W.J. and Watson J.W. (eds), Encyclopedia of Scientific Dating Methods. Springer-Verlag, 285-296.

Gleadow, A.J.W., Belton, D.X., Kohn, B.P., and Brown, R.W. (2002) Fission Track Dating of Phosphate Minerals and the Thermochronology of Apatite. Reviews in Mineralogy and Geochemistry, 48, 579-630.

Gleadow, A.J.W., Duddy, I.R., Green, P.F., and Lovering, J.F. (1986) Confined fission track lengths in apatite: a diagnostic tool for thermal history analysis. Contributions to Mineralogy and Petrology, 94, 405-415.

Gleadow, A.J.W., Gleadow, S.J., Belton, D.X., Kohn, B.P., Krochmal, M.S., and Brown. R.W. (2009a) Coincidence mapping - a key strategy for the automatic counting of fission tracks in natural minerals. Special Publication - Geological Society of London, 324, 25-36.

Gleadow, A.J.W., Gleadow, S.J., Frei, S., Kohlmann, F., and Kohn, B.P. (2009b) Automated analytical techniques for fission track thermochronology. 2009 Goldschmidt Conference Abstracts, A441.

Gleadow, A.J.W., Harrison T.M., Kohn, B.P., Lugo-Zazueta R., Phillips, D. (2015) The Fish Canyon Tuff: A new Look at an Old Low-Temperature thermochronology standard. Earth and Planetary Science Letters, 424, 95-108.

Green, P.F. (1981) “Track-in-track” length measurements in annealed apatites. Nuclear Tracks, 5, 121-128.

588 Green, P.F. (1988) The relationship between track shortening and fission track age reduction
589 in apatite: combined influences of inherent instability, annealing anisotropy, length bias
590 and system calibration. Earth and Planetary Science Letters, 89, 335-352.
591 Green, P.F., Duddy, I.R., Gleadow, A.J.W., Tingate, P.R., and Laslett, G.M. (1986) Thermal
592 annealing of fission tracks in apatite 1. A qualitative description. Chemical Geology.
593 Isotope Geoscience Section, 59, 237-253.
594 Jonckheere, R. and Ratschbacher, L. (2010) On measurements of non-horizontal confined
595 fission tracks (Abstract). In R.W. Brown, Ed. 12th International Conference on
596 Thermochronology, Glasgow, p. 127
597 Ketcham, R.A. (2005) Forward and Inverse Modeling of Low-Temperature
598 Thermochronometry Data. Reviews in Mineralogy and Geochemistry, 58, 275-314.
599 Ketcham, R.A., Donelick, R.A., and Carlson, W.D. (1999) Variability of apatite fission track
600 annealing kinetics: Ⅲ. Extrapolation to geological time scale. American Mineralogist,
601 84, 1235-1255.
602 Ketcham, R.A., Donelick, R.A., Balestrieri, M.L., and Zattin, M. (2009) Reproducibility of
603 apatite fission track length data and thermal history reconstruction. Earth and Planetary
604 Science Letters, 284, 504-515.
605 Ketcham, R.A., Carter, A., and Hurford, A.J. (2015) Inter-laboratory comparison of fission
606 track confined length and etch figure measurements in apatite. American Mineralogist,
607 100, 1452-1468.
608 Kohn, B.P., Gleadow, A.J.W., Brown, R.W., Gallagher, K., O'Sullivan, P.B., and Foster, D.A.
609 (2002) Shaping the Australian crust over the last 300 million years: insights from fission

610 track thermotectonic imaging and denudation studies of key terranes. Australian Journal
611 of Earth Sciences, 49, 697-717.
612 Lal, D., Rajan, R.S., Tamhane, A.S. (1969) Chemical composition of nuclei of Z>22 in
613 cosmic rays using meteoritic minerals as detectors. Nature, 221, 33-37.
614 Laslett, G.M., and Galbraith, R.F. (1996) Statistical modeling of thermal annealing of fission
615 tracks in apatite. Geochimica et Cosmochimica Acta, 60, 5117-5131.
616 Laslett, G.M., Kendall, W.S., Gleadow, A.J.W., and Duddy, I.R. (1982) Bias in measurement
617 of fission track length distributions. Nuclear Tracks, 6, 79-85.
618 Laslett, G.M., Gleadow, A.J.W., and Duddy, I.R. (1984) The relationship between fission
619 track length and track density in apatite. Nuclear Tracks and Radiation Measurements, 9,
620 29-38.
621 Laslett, G.M., Green, P.F., Duddy, I.R., and Gleadow, A.J.W. (1987) Thermal annealing of
622 fission tracks in apatite 2. A quantitative analysis. Chemical Geology: Isotope
623 Geoscience Section, 65, 1-13.
624 Petford, N., and Miller, J.A., (1992) Three dimensional imaging of fission tracks using
625 confocal SLM. American Mineralogist, 77, 529-533.
626 Petford, N., and Miller, J.A. (1993) The study of fission track and other crystalline defects
627 using confocal scanning laser microscopy. Journal of Microscopy, 170, 201-212.
628

Table 1. Annealing experiment details – Fish Canyon Tuff (FCT) apatites

Slide	Primary fluence (n/cm^2)	Annealing conditions	Secondary fluence (n/cm^2)
Control 1 (~16µm)	9×10^{15}	unannealed	
*Control 2 (~13µm)**	9×10^{15}	300°C for 60 mins	
Control 3 (~11µm)	9×10^{15}	350°C for 60 mins	
Control 4 (~8µm)	9×10^{15}	370°C for 60 mins	
FCT A (4:1 13µm, 16µm)	9×10^{15}	300°C for 60 mins	2.0×10^{15}
FCT B (4:1 11µm, 16µm)	9×10^{15}	350°C for 60 mins	1.5×10^{15}
FCT C (1:4 11µm, 16µm)	9×10^{15}	350°C for 60 mins	2.5×10^{16}
FCT E (1:1 13µm, 16µm)	9×10^{15}	300°C for 60 mins	8.0×10^{15}
FCT F (1:4 8µm,16µm)	9×10^{15}	370°C for 60 mins	1.7×10^{16}
FCT G (4:1 8µm, 16µm)	9×10^{15}	370°C for 60 mins	1.1×10^{15}
FCT H (1:1 8µm, 16µm)	9×10^{15}	370°C for 60 mins	4.3×10^{15}

Controls 1-4 are single-irradiated samples representing three of the four discrete length components in the FCT mixtures. The remaining seven FCT samples contain two component mixtures of tracks following the second irradiation. Brackets show the ratio of the two components and their respective mean lengths in each case.
*Control 2 was not available for this study.

Table 2. Mean track length data for all samples

Sample ID	Analyst 1 (UoM)						Analyst A (UCL)		
	N	HCT_p (SE) (µm)	SD (µm)	N	ACT_t (SE) (µm)	SD (µm)	N	HCT_p (SE) (µm)	SD (µm)
Control 1	142	15.85(06)	0.77	170	15.89(05)	0.70	100	16.23(07)	0.66
Control 3	143	10.78(07)	0.90	184	10.77(07)	0.98	100	10.90(09)	0.89
Control 4	139	8.28(21)	2.47	181	8.25(18)	2.36	100	8.48(25)	2.47
FCT A	135	14.43(10)	1.11	191	14.34(09)	1.21	100	14.43(13)	1.27
FCT B	151	12.12(17)	2.14	213	11.99(14)	2.06	100	12.39(23)	2.24
FCT C	153	15.13(17)	2.06	198	15.01(15)	2.18	100	15.07(21)	2.11
FCT E	198	14.92(10)	1.37	255	14.90(08)	1.35	100	14.87(14)	1.44
FCT F	187	14.52(30)	4.14	253	14.34(27)	4.30	100	14.91(29)	2.88
FCT G	73	12.31(54)	4.63	120	12.51(40)	4.37	100	13.53(44)	4.37
FCT H	141	14.59(29)	3.48	209	14.69(23)	3.34	100	14.56(38)	3.75
DUR	162	14.14(07)	0.86	256	14.13(05)	0.86	-	-	-
DUR-4	157	14.20(06)	0.80	169	14.24(06)	0.79	-	-	-

UoM = University of Melbourne; UCL = University College London; N = number of confined fission tracks measured; HCT_p = mean projected length for horizontal confined tracks (dip <10°); ACT_t = mean true 3D length for tracks of all orientations; SE = standard error of the mean; SD = standard deviation.

643

644 **Table 3.** Mean 3D track length data for horizontal ($\theta \leq 10°$) and inclined ($\theta > 10°$) confined tracks

645

Sample ID	HCT length measurements ($\theta \leq 10°$)			ICT length measurements ($\theta > 10°$)			Difference
	N	HCT_t (µm)	SD (µm)	N	ICT_t (µm)	SD (µm)	(µm)
Control 1	143	15.91(06)	0.70	28	15.77(12)	0.66	0.14
Control 3	143	10.81(08)	0.9	41	10.66(23)	1.44	0.15
Control 4	139	8.30(21)	2.48	42	8.08(29)	1.90	0.22
FCT A	135	14.47(10)	1.12	56	14.01(18)	1.35	0.46
FCT B	151	12.16(17)	2.13	62	11.58(23)	1.81	0.57
FCT C	153	15.18(17)	2.06	45	14.42(37)	2.46	0.76
FCT E	198	14.97(10)	1.37	57	14.66(17)	1.27	0.31
FCT F	187	14.56(30)	4.15	66	13.73(58)	4.67	0.82
FCT G	73	12.36(54)	4.65	47	12.73(57)	3.92	-0.37
FCT H	141	14.64(29)	3.49	68	14.78(37)	3.02	-0.13
DUR	162	14.20(07)	0.86	94	14.01(09)	0.86	0.19

646 HCT_t = mean 3D length of Horizontal Confined Tracks; ICT_t = mean 3D length of Inclined Confined

647 Tracks. Brackets show the standard error of the mean.

648

649

650

651 **Table 4.** Mean 3D track lengths for DUR measured using different vertical step-sizes

652

Step-size (µm)*	N	ACT_t (SE) (µm)	SD (µm)
0.1	80	14.08(09)	0.87
0.2	80	14.14(08)	0.89
0.3	80	14.10(08)	0.87

653 * Distance between captured planes in the image stack.

654

655

Table 5. Comparison of conventional and confocal laser scanning microscopy measurements

Sample	N	Conventional wide field microscopy		Confocal laser scanning microscopy		Difference	Dip angle* (θ°)	
		ACT_t (SE) (μm)	SD (μm)	ACT_t (SE) (μm)	SD(μm)	(μm)	Mean	Maximum
Control 1	94	16.00(08)	0.80	16.12(09)	0.86	0.12	3.17	13.5
Control 3	96	10.83(09)	0.85	11.04(09)	0.84	0.21	2.37	9.9
Control 4	69	8.19(31)	2.57	8.41(30)	2.52	0.22	2.36	13.1
FCT A	73	14.41(13)	1.14	14.54(13)	1.15	0.13	2.58	10.7
FCT B	92	12.03(21)	2.01	12.20(21)	2.04	0.17	2.83	15.7
FCT C	61	15.02(28)	2.16	14.92(28)	2.18	-0.10	3.55	11.3
FCT E	125	14.86(13)	1.43	14.96(13)	1.49	0.10	3.12	12.3
FCT F	106	15.43(30)	3.10	15.43(30)	3.11	0.00	2.82	11.7
FCT G	51	13.54(58)	4.16	13.63(59)	4.19	0.09	4.05	13.3
FCT H	82	15.23(30)	2.71	15.19(31)	2.78	-0.04	3.41	12.8

*Dips determined from conventional wide field microscopy. All measurements are mean track lengths of measurements on the same tracks imaged by both methods.

Table 6. Replicate measurement of mean track lengths by Analyst 1 after a 2.5-year interval

Sample	1st Analysis			2nd Analysis			Difference
	N	ACT_t (SE) (μm)	SD (μm)	N	ACT_t (SE) (μm)	SD (μm)	(μm)
Control 1	171	15.91(06)	0.76	176	15.88(06)	0.78	0.03
FCT A	191	14.34(09)	1.21	192	14.22(09)	1.23	0.12
FCT B	213	11.99(14)	2.06	213	11.92(14)	2.10	0.07
FCT G	120	12.51(40)	4.37	120	12.42(41)	4.49	0.09
DUR	256	14.13(05)	0.86	238	14.08(06)	0.86	0.05

Measurements made on the same captured image sets.

668 Table 7. Mean track length measurements for DUR-4 by six University of Melbourne analysts

669

Analyst	N	ACT_t (SE) (μm)	SD (μm)
1	169	14.24(06)	0.79
2	120	14.20(07)	0.80
3	122	14.29(09)	0.98
4	120	14.12(07)	0.82
5	130	14.14(07)	0.84
6	100	14.23(09)	0.94

670

671

672 **Figure Captions**

673 **Figure 1.** Three different measures of fission track length (colored lines) shown against a 3D
674 coordinate system. These show the true track length l_t (red), the projected length l_p (green),
675 and the apparent length (l_a, uncorrected for refraction) in apatite. Apparent (measured) and
676 true depths to the end of the track are represented by d_a (red), and d (blue) respectively, these
677 being related by the refractive index. Angles shown are the azimuth to the *c*-axis ω, the true
678 angle to the *c*-axis φ, and the dip angle θ. XOY is a plane parallel to the observation surface
679 and contains one end of a confined fission track at the origin O. The X-axis is parallel to the
680 crystallographic *c*-axis (modified after Galbraith and Laslett, 1988).

681

682 **Figure 2.** Confined fission track length distributions for all samples showing the 'true' 3D
683 lengths (i.e. corrected for dip and refraction), l_t, as grey histograms and red relative
684 probability plots. Two sub-sets of these 3D length data are also shown for 'horizontal' (blue)
685 ($\theta \leq 10°$, blue), and 'inclined' ($\theta > 10°$, green) confined tracks as relative probability curves.
686 Numbers in colors for each sample denote mean confined track lengths (ACT = All, HCT =
687 Horizontal, and ICT = Inclined) and standard deviations for the three distributions. The
688 number of tracks measured in each case is shown in brackets.

689

690 **Figure 3.** Residual differences between mean confined track lengths determined for
691 'horizontal' ($\theta \leq 10°$, HCT), inclined ($\theta > 10°$, ICT), and all (ACT) fission tracks. Subscripts
692 denote projected (l_p) and true (l_t) lengths as shown in Fig. 1. *Measured by Analyst A at
693 UCL, all other measurements were made by Analyst 1 at UoM.

Figure 4. (A) 3D lengths of individual confined tracks plotted against dip angle, θ, in Control, FCT and DUR apatite samples. Alignment of the data into discrete dot-curve arrays is a result of the depth measurements being limited to specific image planes in the z-stack (Fig. 5). The vertical image spacing was 0.3 µm for all samples except DUR, for which it was 0.2 µm. The different length components can be clearly seen in the mixtures of unannealed and moderately to highly annealed tracks (FCT B, C, F, G, H). (B) Histogram of dip angles for all tracks measured showing that the number of observed tracks decreases rapidly with increasing dip angle θ, and that almost no confined tracks are observed at dips greater than 30°.

Figure 5. (A) Individual 3D fission track lengths plotted against dip angle in sample FCT C, measured in a transmitted light z-stack with a vertical image spacing of 0.3 µm. Colors show the depth component of each 3D length measurement and how the results align into discrete dot-curves corresponding to track lengths that cover the same fixed intervals between the image planes defining the ends of each track. (B) Simplified diagram illustrating how dip angles can increase continuously while the depth is limited to discrete intervals between two image planes. Dashed lines represent two planes in the image stacks where the track ends are located.

Figure 6. (A) Variation in lengths of individual confined tracks as a function of dip angle in apatite DUR, as summarized by the mean lengths in Table 4. Results are shown for measurements made on z-stack images with three different step-sizes between the image

716 planes. The spacing between the dot-curves varies according to the step-size, but in all three
717 cases the overall range of variation is the same. (B) Comparison of paired 3D track lengths
718 for 88 individual confined tracks measured in image stacks with step-sizes of 0.3 μm and
719 0.1μm.
720
721 **Figure 7.** Comparison of images on the same horizontal track (dip ≈ 0°) captured by (A)
722 confocal laser scanning microscopy (l_t = 11.03 μm) and (B) conventional wide field
723 transmitted light microscopy (l_t = 10.93 μm). Confocal imaging increases the resolution
724 enabling the ends of the track to be more clearly defined, but is only useful for relatively
725 shallow dipping tracks.
726
727 **Figure 8.** Comparison of individual 3D track lengths on the same confined tracks measured
728 by confocal laser scanning microscopy and conventional wide field transmitted light
729 microscopy for ten samples studied, confirming a high degree of consistency between the two
730 methods.
731
732 **Figure 9.** Mean confined track lengths and uncertainties for apatite sample DUR-4. (A)
733 Summary of all track length results from 55 analysts in the inter-laboratory comparison of
734 Ketcham et al., (2015) measured as projected lengths of horizontal confined tracks (HCT_p).
735 Data shown in red are the mean 3D lengths (ACT_t) measured by six UoM analysts in this
736 study (Table 7). Green dashed lines denote the 14.12 ± 0.08 μm mean value reported by
737 Ketcham et al., (2015). (B) Expansion of the mean track length results from this study (Table

738 7). Blue dashed lines denote the mean value of 14.20 ± 0.03 μm for the six measurements.

739

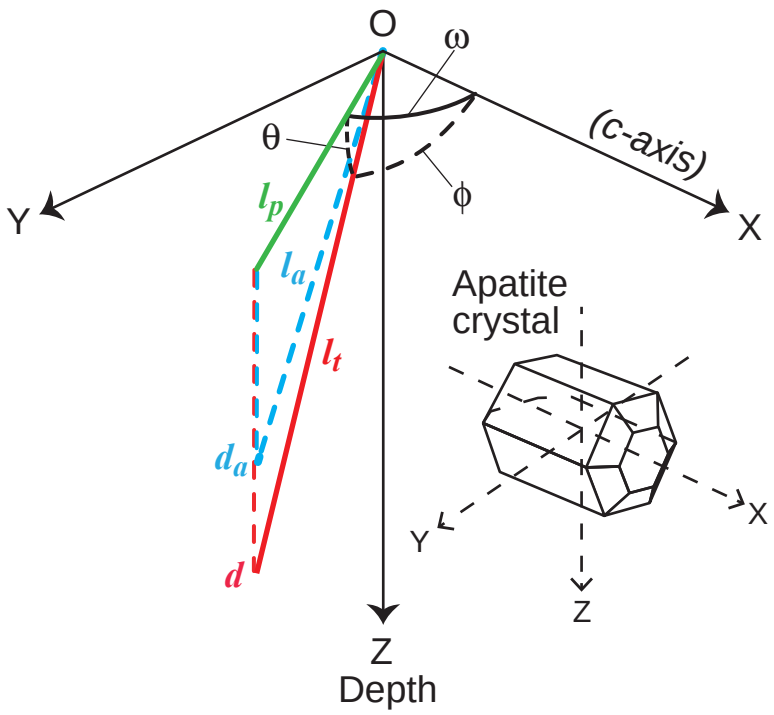

Figure 1 revised

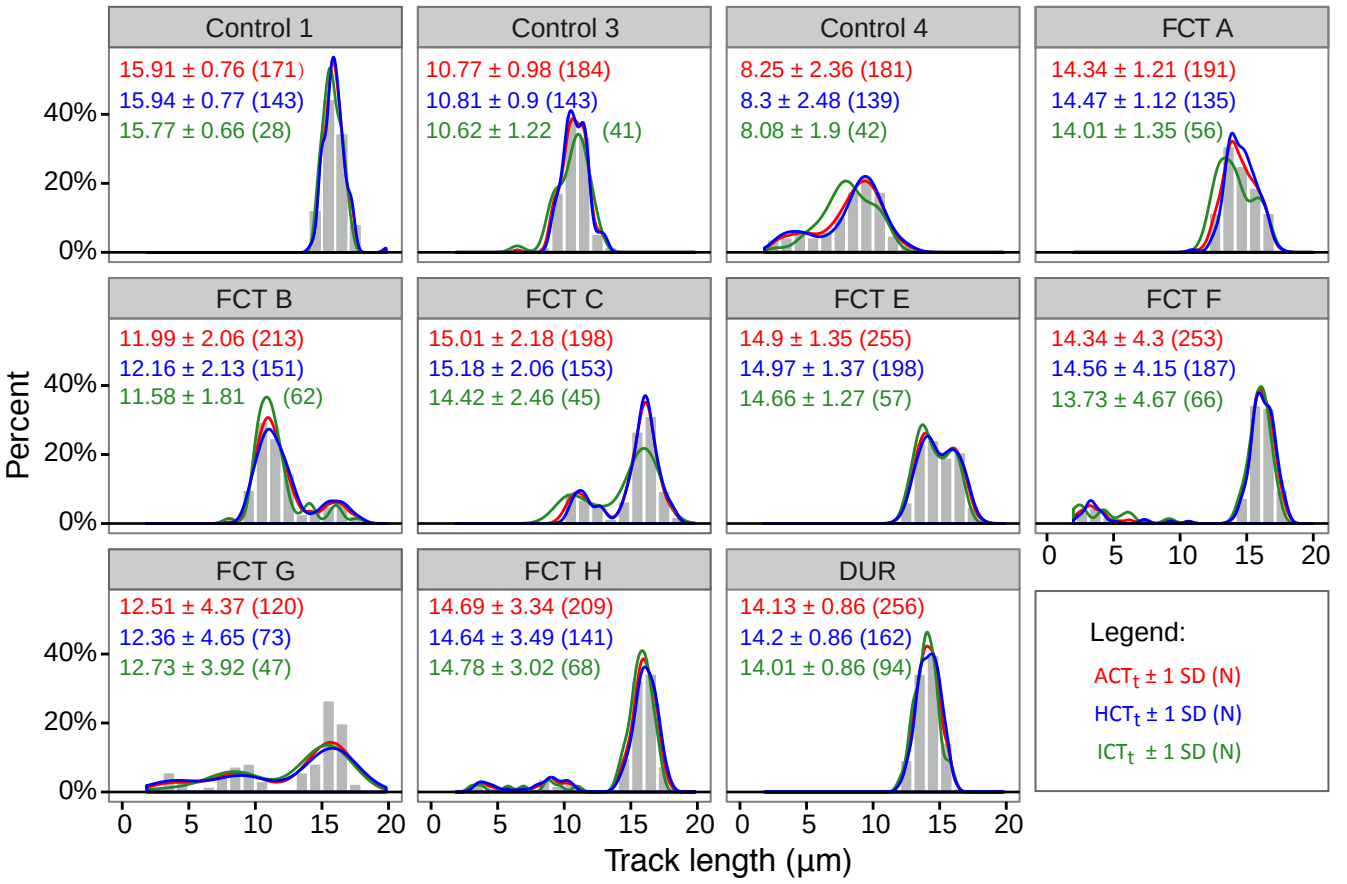

Figure 2 revised

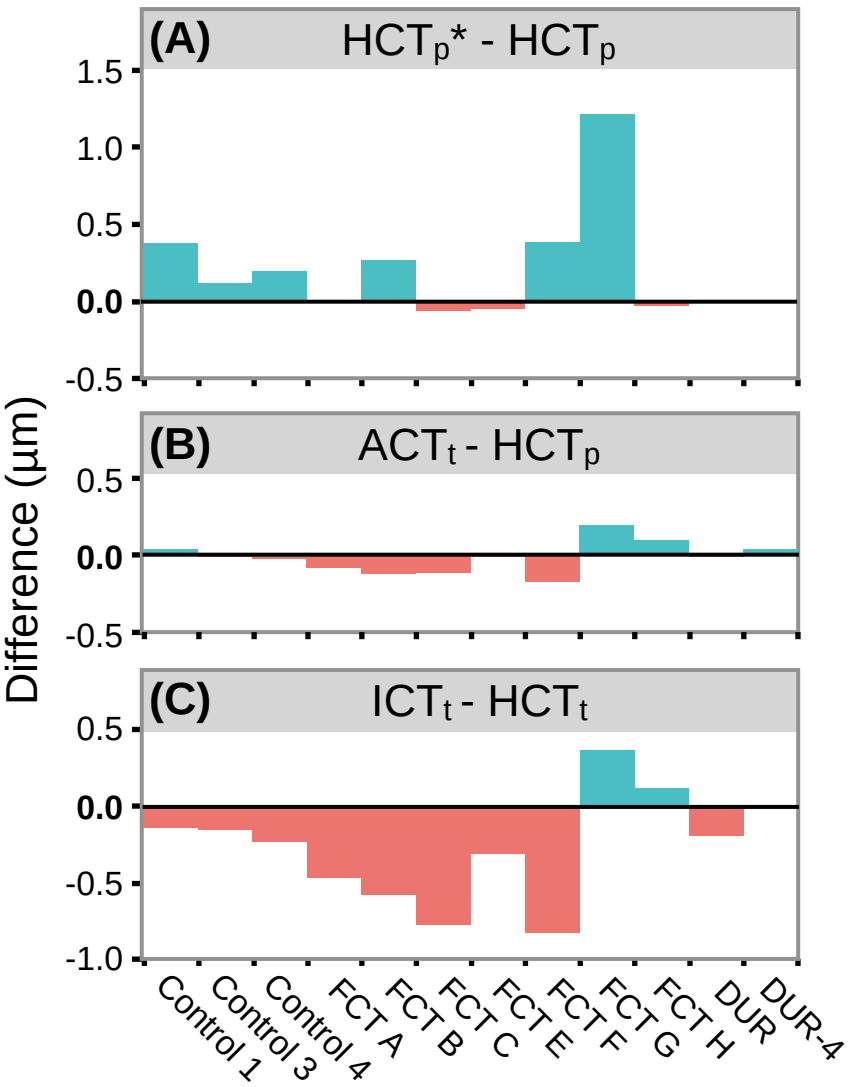

Figure 3 Revised

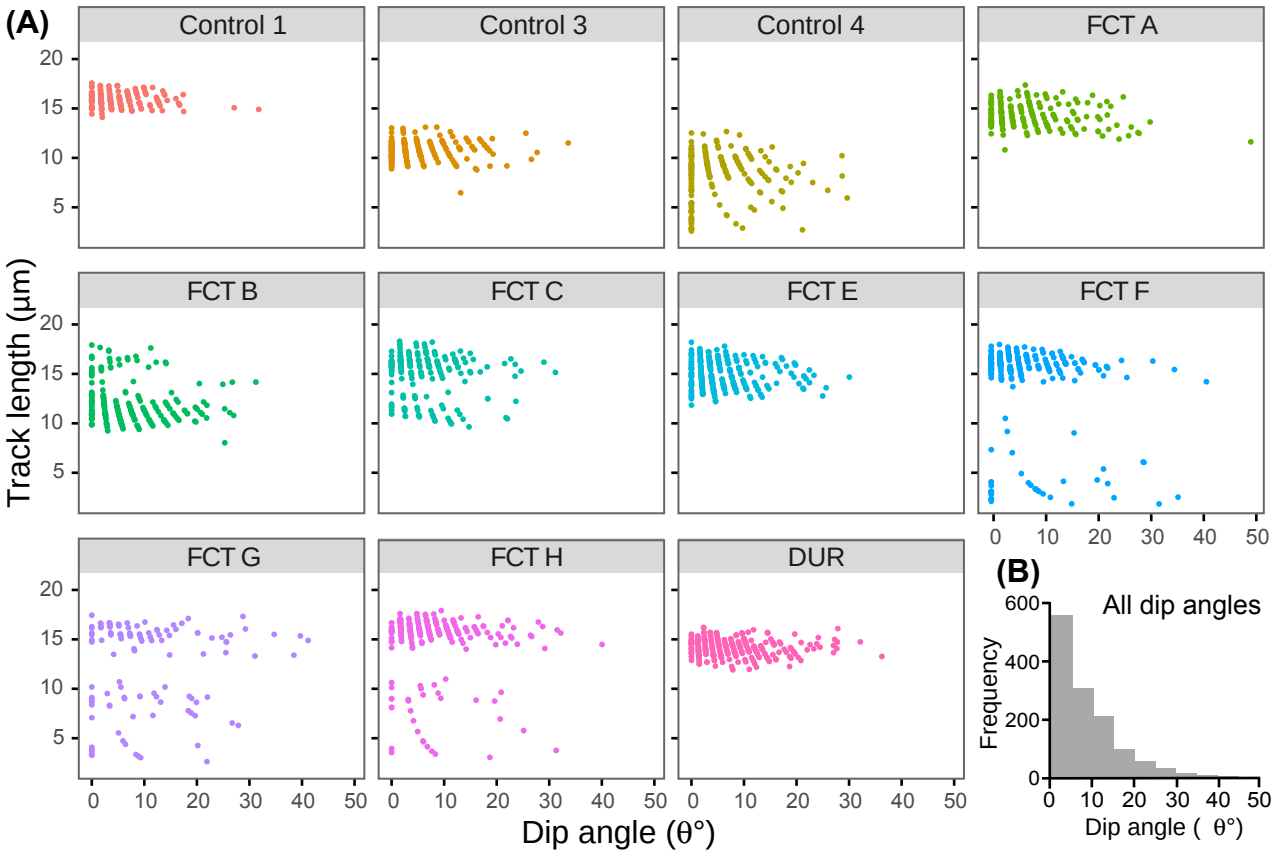

Figure 4 revised

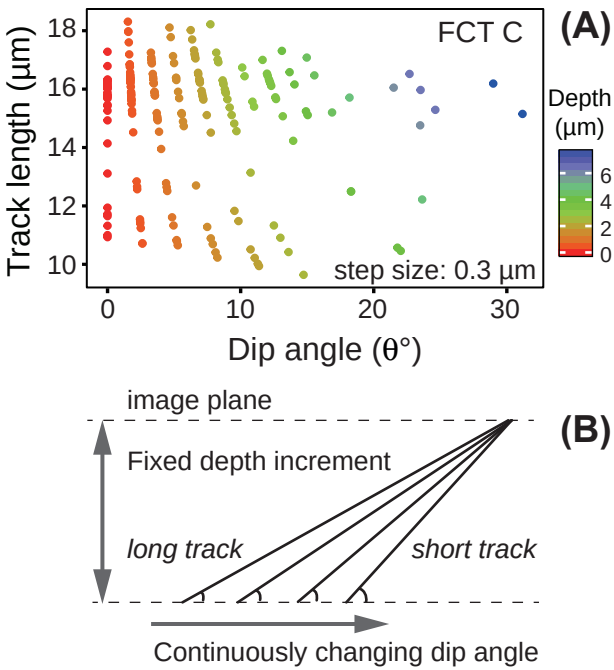

Figure 5 revised

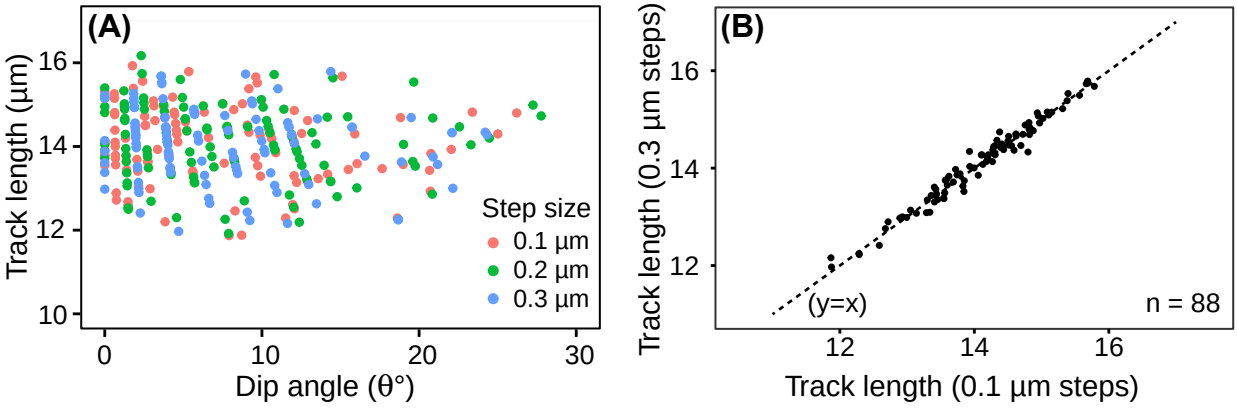

Figure 6 revised

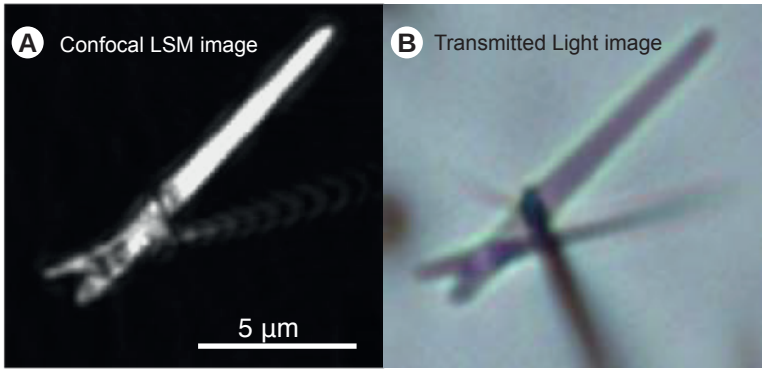

Figure 7 revised

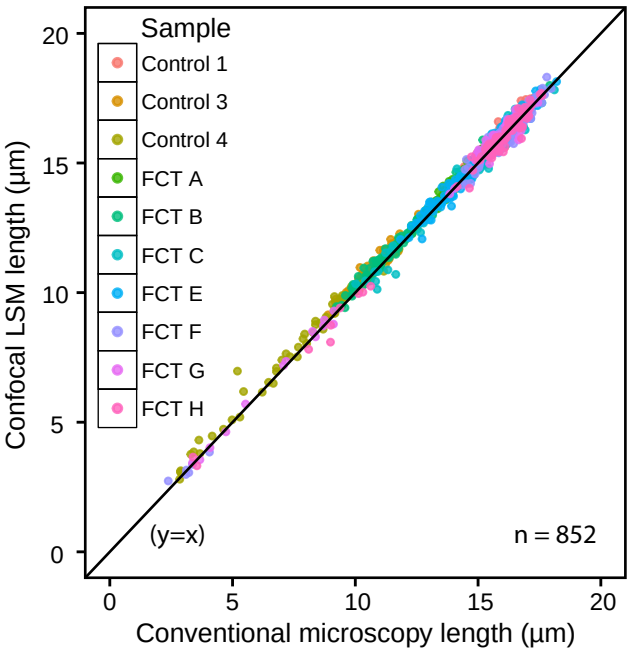

Figure 8 revised

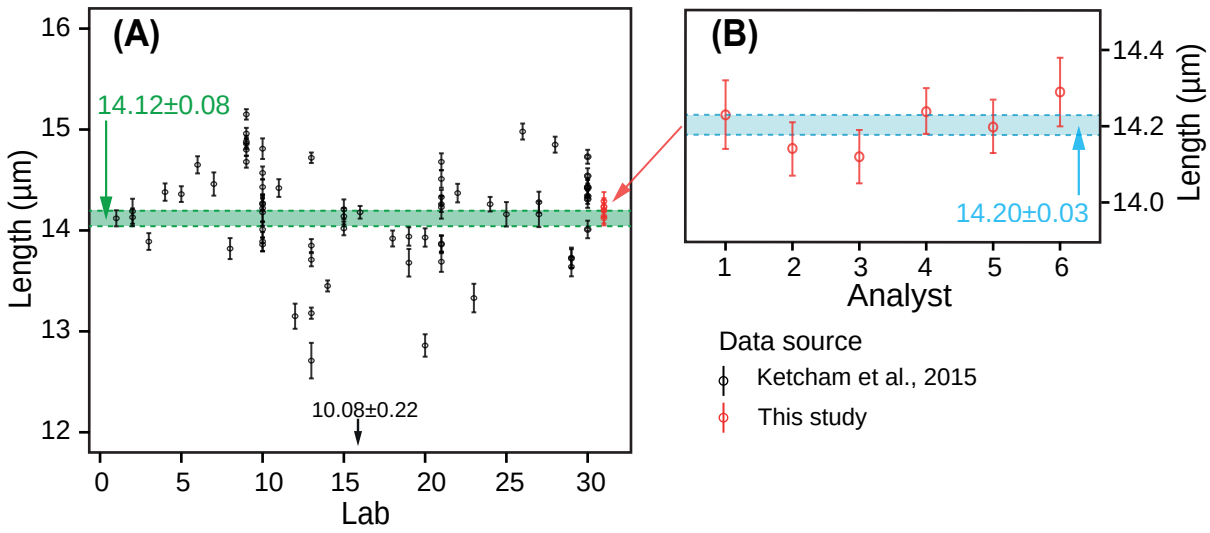

Figure 9 revised